SCREENING CHEMICALS FOR ESTROGEN RECEPTOR BIOACTIVITY USING A COMPUTATIONAL MODEL

Patience Browne[1*], Richard S. Judson[2], Warren Casey[3], Nicole Kleinstreuer[4], Russell S. Thomas[2]

[1]US EPA, Office of Chemical Safety and Pollution Prevention, Washington DC; [2]US EPA, Office of Research and Development, Research Triangle Park, NC; [3]National Toxicology Program, Interagency Center for the Evaluation of Alternative Toxicological Methods, Research Triangle Park, NC; [4]ILS/NICEATM, Research Triangle Park, NC

The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency.

*Address correspondence to P. Browne, US EPA, Office of Science Coordination and Policy, 1200 Pennsylvania Ave NW (7203M), Washington, DC 20460, Telephone: 202-564-8474, E-mail: browne.patience@epa.gov.

Abstract

The U.S. Environmental Protection Agency (EPA) is beginning to use high-throughput and computational methods for regulatory applications in the Endocrine Disruptor Screening Program (EDSP). To use these new tools for regulatory decision making, computational methods must be appropriately validated. Traditional validations of toxicity tests are time intensive, evaluate a relatively small number of chemicals, and are not well-suited to high-throughput methods. Here we describe a multi-step, performance-based validation establishing scientific confidence in new computational methods and demonstrating these tools are sufficiently robust to be used in a regulatory context. Results from 18 estrogen receptor (ER) ToxCast high-throughput screening assays, measuring different points along the signaling pathway with different assay technologies, were integrated into a computational model. The resulting ToxCast ER model scores range from 0 (no activity) to 1 (bioactivity of the native ligand, 17β-estradiol) and can discriminate ER bioactivity from assay-specific interference and cytotoxicity. ToxCast ER model performance was evaluated for 40 *in vitro* and 43 *in vivo* reference chemicals. ToxCast ER model results were also compared to EDSP Tier 1 screening assays in current regulatory practice for a diverse set of more than 100 chemicals. ToxCast ER model accuracy was 95% when compared to the large set of *in vitro* and *in vivo* reference chemicals. In addition, the ToxCast ER model predicted the outcomes of EDSP Tier 1 guideline and other uterotrophic studies with > 90% accuracy. The performance of the high-throughput assays and ToxCast ER model predictions of agonist bioactivity demonstrates these methods are sensitive, specific, quantitative, and efficient; and thus protective of human health and the environment. EPA is accepting ToxCast ER model data for over 1800 chemicals as alternatives for the EDSP Tier 1 ER binding, ER transactivation, and uterotrophic assays. The use of high-throughput and computational methods will dramatically increase EPA's ability to rapidly screen chemicals for endocrine bioactivity, and are an alternative to animal-based EDSP Tier 1 ER binding and uterotrophic assays. The application of these alternative, innovative tools for screening chemicals for endocrine bioactivity represents the first step in a paradigm shift for chemical safety testing, and the first systematic application of ToxCast data in an EPA regulatory program.

Introduction

The US Environmental Protection Agency's (EPA) Endocrine Disruptor Screening Program (EDSP) was established in 1999 for the purpose evaluating potential risk of endocrine disruption in humans and wildlife from exposure to pesticide chemicals and drinking water contaminants. To screen chemicals for potential estrogen, androgen, and thyroid bioactivities, EPA developed a battery of five *in vitro* and six *in vivo* Tier 1 screening assays[1] to

evaluate potential endocrine bioactivity. To test for endocrine disruption, EPA developed *in vivo* multigenerational Tier 2 tests that include apical endpoints to identify adverse effects, and establish quantitative dose response relationships[2]. In 2009, EPA published a final list of 67 pesticide chemicals (List 1), and in 2011 issued EDSP Tier 1 test orders on these chemicals[3]. Fifteen List 1 chemicals were voluntarily withdrawn from the pesticide market. EPA is currently reviewing data submitted in response to List 1 Tier 1 test orders, along with other scientifically relevant information, and is developing weight of evidence evaluations of potential endocrine bioactivity with a determination of further testing that may be required for the remaining 52 of the 67 chemicals. A second list of pesticide and high production volume chemicals (List 2) proposed for Tier 1 screening was published in June 2013[4]; however, test orders have yet to be issued. The remaining universe of pesticide chemicals and drinking water contaminants includes approximately 10,000 chemicals to be screened for potential endocrine bioactivity[5].

In response to the US National Academy of Sciences report, Toxicity Testing in the 21st Century[6] and the US President's 2012 directive[7], EPA began a multi-year transition from existing EDSP test methods towards utilizing more rapid, cost-effective computational models and high-throughput assays. The transition to using computational toxicology approaches to prioritize and screen thousands of EDSP chemicals has been outlined by the Agency in two strategic planning documents[8, 9]. However, to use new computational toxicology approaches in the existing EDSP screening and testing framework, they must be validated (*i.e.*, determined to be fit for purpose) and perform as well or better than existing methods currently in practice.

A variety of pesticides and environmental chemicals act as estrogen receptor (ER) agonists[10, 11] and, while the scope was expanded to consider androgen and thyroid active environmental chemicals, the EDSP was originally established in response to statutory mandates in the Federal Food Quality Protection and Safe Water Drinking Acts compelling EPA to evaluate potential xenoestrogens. As a result, the Tier 1 screening battery was weighted towards assays that detect potential ER interactions. Because EDSP Tier 1 assay endpoints that measure interaction with the ER only measure agonism and environmental chemicals that act through the ER are primarily expected to act as agonists, we focused this demonstration using computational toxicology tools to evaluate ER agonist bioactivity.

In this manuscript, we present a fit-for-purpose, performance-based approach to validate computational toxicology tools for contributing to the weight of evidence evaluation of a chemical's potential ER bioactivity and as alternative data for specific Tier 1 endpoints intended to identify *in vitro* and *in vivo* ER interactions. Data from high-throughput *in vitro* screening assays included in EPA's ToxCast program[12] were integrated into a computational network model of ER pathway activity[13], and model performance was compared with *in vitro* and *in vivo* reference chemicals identified from a semi-automated curation of peer-reviewed endocrine toxicology literature[14]. A variety of publications have previously described the ToxCast program[15, 16], validation of the high-throughput screening assays[17], and endocrine assays[18, 19]. However, this paper is the first description of an approach for using computational methods to set priorities for chemical screening and testing and to use high-throughput screening data as an alternative for regulatory guideline studies. This approach is consistent with recommendation of EPA's strategic plan for evaluating the toxicity of chemicals[8] and the 2007 NAS report[6], specifically to: 1) provide broad coverage of chemicals examined; 2) reduce the cost and time of toxicity testing; 3) reduce animal use; and 4) develop a robust scientific basis for assessing health effects of environmental agents.

Methods

In Vitro *Assays*

Details of the *in vitro* assays are described on EPA's ToxCast website[12] and in a variety of publications[20, 21]. Briefly, potential estrogen receptor bioactivity was measured in 18 high-throughput *in vitro* assays run in EPA's ToxCast program. The ER pathway assays (Table 1) include three cell-free biochemical radioligand ER binding assays (NVS_NR_bER, NVS_NR_hER, NVS_NR_mERa[22, 23]); three protein complementation assays that measure formation of ER dimers and test for activity against ERα and ERβ, each measured at two time points (OT_ER_ERaERa, OT_ER_ERaERb, OT_ER_ERbERb); an assay measuring interaction of the mature transcription factor with DNA at two time points (OT_ERa_EREGFP); two reporter gene assays measuring RNA transcript levels (ATG_ERa_TRANS_up, ATG_ERE_CIS_up[24]); two assays measuring reporter protein level readouts (Tox21_ERa_BLA_Agonist_ratio, Tox21_ERa_LUC_BG1_Agonist[25]); an ER-sensitive cell proliferation assay (ACEA_T47D[26]); and two transactivation antagonist assays (Tox21_ERa_BLA_Antagonist_ratio, Tox21_ERa_LUC_BG1_Antagonist[25]).

This combination of biochemical and cell-based *in vitro* assays relies on different technologies and probes different points in the ER signaling pathway (Table 1). Though the assays are primarily human proteins and/or cell types, the suite of 18 assays include human, murine, and bovine ER binding assays and ER pathway interactions in a variety of human tissue types (Table 1). Every *in vitro* assay is potentially subject to technology-specific interference (*e.g.*, chemicals that interfere with the receptor protein, are luminescent, are cytotoxic, *etc.*). However, combining data from multiple orthogonal assays and integrating data in a network model of the entire ER pathway allows for detection of false positives and a more confident assessment of the "true" *in vitro* estrogenic bioactivity of the tested chemical.

Concentration Response Analysis and Computational Modeling

Chemicals were run in concentration-response format in all assays except the cell-free binding assays. Cell-free binding assays were initially run at a single concentration (25 μM), and if activity was detected, the assay was run in concentration-response format. Concentration-response data from the *in vitro* assays were fit to a series of three models that included a four parameter Hill model, a modified Hill model with gain-loss at high concentrations, or a constant (no concentration-response) model[27]. The best model was statistically selected using the Aikake Information Criteria value. All concentration-response data were analyzed using the ToxCast data analysis pipeline, which automates the processes of baseline correction, normalization, curve-fitting, hit-calling and detection of a variety of potential confounders[28]. To allow computational synthesis of the different *in vitro* assays, concentration-response curves were generated for each assay for a series of 14 concentrations from 0.01-100 μM.

The concentration-response curves for all 18 assays were included in a computational network model, referred to here as the ToxCast ER model for bioactivity and described in Judson *et al.*[13] and EPA[29]. The computational model integrates data from the 18 *in vitro* assays measuring ER agonist and antagonist response in an unweighted manner, while subtracting background and other non-specific assay interference including cytotoxicity. Model output includes separate agonist and antagonist area under the curve (AUC) scores that range from 0 to 1.0. All outputs were normalized to the agonist response of 17β-estradiol. Model scores were truncated at values < 0.001, considered to have no ER bioactivity and given 0 scores, as a value <0.001 implies an AC_{50} greater than 10 millimolar which is several orders of magnitude greater than the highest concentrations tested in ToxCast assays. ToxCast ER agonist scores $\geq$ 0.1 were considered positive; model score of 0.1 equates to an AC_{50} of about 100 μM and approximates the limit of detection. Model scores of 0.1> AUC >0 for either

agonist or antagonist activity were considered equivocal because values in this range were associated with bioactivity occurring at concentrations near the upper testing limit of the high-throughput assays.

Performance-based Computational Model Validation

To assess the utility and limitations of the ToxCast ER model, we adopted a performance-based validation approach consistent with the Organization for Economic and Commercial Development (OECD) conceptual framework for testing and assessment of potential endocrine disrupting chemicals[30]. In principle, this method can be used to assess the applicability of any test method or set of methods that meets defined performance standards. The ToxCast ER model was validated using three separate approaches: 1) *in vitro* reference chemicals; 2) literature-derived *in vivo* studies that were methodologically similar to the Tier 1 uterotrophic bioassay (*i.e.*, "guideline-like"), including a subset of chemicals with independently confirmed activities in at least two studies and considered *in vivo* reference chemicals; and 3) results for the List 1 chemicals in the EDSP Tier 1 assays. For the *in vitro* reference chemical validation, 40 chemicals (28 agonists of differing potencies and 12 inactive chemicals) were selected for reproducible agonist results in *in vitro* ER binding and transactivation assays, and to include a diverse set of chemical structures (Table 2)[31, 32]. All *in vitro* reference chemicals were run in the 18 high-throughput ToxCast ER assays and the resulting ToxCast ER model scores for agonist bioactivity were compared with results anticipated from low or medium throughput *in vitro* assays[31].

In vivo reference chemicals were established from a literature search of short-term rodent uterotrophic assays that were methodologically similar to the OECD[33] and EDSP Tier 1 battery uterotrophic[34] assays. A comprehensive search and review of uterotrophic studies published in peer-reviewed literature was performed as previously described[14]. Briefly, the chemical name and chemical abstract services registry number (CASRN) were used to search PubMed, the EPA's Aggregated Computational Toxicology Resource (ACToR)[35], and the US Food and Drug Administration's Endocrine Disruptor Knowledge Base (EDKB)[36] for the ~1800 chemicals run in the ToxCast *in vitro* ER assays with "uterotrophic assay", "uterotrophic", "uterotropic", and "uterine weight" as modifier terms. The articles identified were reviewed for methodological consistency with the EDSP uterotrophic assay guidelines[34] based on: 1) age and species of animals used (immature rat or ovariectomized mouse or rat); 2) number of animals per treatment group; 3) number of treatment groups; 4) route of chemical administration; 5) length of dosing; and 6) time of necropsy. Over 1000 articles were identified, entered into a database, and independently reviewed by two scientists. Only studies that met all six minimum criteria were considered "guideline-like" and used in this analysis.

Chemical data from curated guideline-like uterotrophic studies were considered with two levels of stringency. First, a chemical was considered positive for potential *in vivo* ER agonist bioactivity if a significant increase in uterine weight among treated animals was reported, and negative if no significant increase in uterine weight was reported in any guideline-like study. Second, only chemicals tested in two or more guideline-like uterotrophic studies were used. Chemicals that resulted in a significant increase in uterine weight in two or more independent guideline-like studies were considered positive, while those chemicals that showed negative results in all studies (two or more) were considered negative. This second subset of chemicals were referred to as *in vivo* reference chemicals since study results were reproducible (Table 3).

For comparisons with EPA's List 1 chemicals where both ToxCast and Tier 1 data were available, ToxCast ER model agonist scores were compared with results of EDSP Tier 1 assays that indicate potential interaction of a test chemical with the estrogen receptor. The EDSP Tier 1 guideline *in vitro* ER binding assay[37] uses primarily ERα obtained from rat uterine cytosol and does not distinguish between ER agonist and antagonist activities. The competitive binding assay measures test chemical displacement of radioligand ([^{3}H]17β-estradiol) from the

ER across a range of concentrations in three independent runs. Results of the assay are "positive" if the test chemical displaces >50% of radioligand (and Log(IC_{50}) is calculated), "equivocal" if test chemical displaces <50% but >25% of radioligand, and "negative" if test chemical displaces <25% of radioligand. The Tier 1 *in vitro* Estrogen Receptor Transcriptional Activation Assay[38] measures chemiluminescence in response to an ERα-mediated increase in luciferase gene expression (*i.e.,* agonist activity). A test chemical is "positive" if the maximum response induced by the test chemical is ≥ 10% of the maximal response (RPC_{max} [positive control: 17β-estradiol]) in at least two of three assay runs (*i.e.,* RPC_{max} ≥10). If the test chemical fails to achieve at least 10% of the response of the positive control, a negative response is recorded for the test chemical. The Tier 1 uterotrophic assay[34] is a short-term, *in vivo* assay designed to detect exogenous estrogen agonist activity indicated by an increase in uterine weight in animals in which the hypothalamic-pituitary-ovarian axis is not functional.

Application of ToxCast ER model to EDSP Chemicals

ER model agonist scores were examined for the ~1800 EDSP chemicals evaluated in ToxCast. These chemicals include some List 1 pesticides for which EPA has issued Tier 1 test orders and List 2 chemicals (both pesticides and drinking water contaminants) identified by EPA as candidates to receive the next group of Tier 1 test orders.

Results

Concentration-response curves of the 18 high-throughput *in vitro* ER assays were integrated into an orthogonal network model of ER pathway bioactivity[13]. Model outputs include an integrated measure of agonist bioactivity, antagonist bioactivity, as well as "false positive" signaling due to cytotoxicity or technology-specific interference[13]. For reasons described previously in this paper, ER agonist bioactivity was considered in these analyses.

The performance-based assessment of the ToxCast ER model detecting agonist bioactivity relied on sets of *in vitro* reference chemicals, *in vivo* reference chemicals, guideline-like uterotrophic studies, and results of EDSP Tier 1 assays (Figure 1). Positive ToxCast ER agonist scores ≥ 0.1 are approximately equivalent to an AC_{50} of about 100 μM, approaching the limit of detection of detection for most assays. Model scores < 0.1 and >0 were associated with bioactivity near the highest concentrations test and were considered equivocal. Among the 40 *in vitro* agonist reference chemicals, 38 showed unequivocal ToxCast ER model bioactivity defined as either AUC = 0 or AUC ≥ 0.1. ToxCast ER model bioactivity was equivocal (0<AUC<0.1) for two chemicals, including one very weak positive chemical, di-n-butyl phthalate, and one inactive chemical, haloperidol (Table 2). These two equivocal chemicals were excluded from further analysis. Of the 38 remaining reference chemicals, 25 of 27 positive reference chemicals had ToxCast ER model agonist scores ≥ 0.1, including all strong, moderate and weak agonists, which were detected with 100% accuracy (Table 2). Two very weak reference chemicals, diethylhexyl phthalate (DEHP) and dicofol, had no detected bioactivity (model score = 0) and were potentially false negatives. All 11 remaining inactive reference chemicals had no ER agonist bioactivity. For the 38 *in vitro* reference chemicals with unequivocal AUC values, the ToxCast ER agonist model scores had an accuracy of 95% (38/40; Table 4).

Performance of the ToxCast ER model agonist bioactivity was also evaluated for *in vivo* reference chemicals with effects that were independently verified in two or more guideline-like uterotrophic studies (Table 3). Among the 43 *in vivo* reference chemicals, 39 had unequivocal ToxCast ER model scores defined as either AUC = 0 or AUC ≥ 0.1, including 30 chemicals with independently confirmed positive uterotrophic responses and nine chemicals with confirmed negative responses (Figure 1). Four chemicals that were inactive in uterotrophic studies (dibutyl

phthalate, dicyclohexyl phthalate, dihexyl phthalate, and fenvalerate) had very low equivocal ToxCast ER model scores (Table 3) and were excluded from further analyses. The potential false positive chemical, kaempferol, was negative in uterotrophic studies but had modest ER agonist model bioactivity (AUC=0.25, Table 3), though the positive result was consistent with other lower throughput *in vitro* ER assays[31, 39]. The potential false negative chemical, octamethylcyclotetrasiloxane (D4), was positive in multiple uterotrophic studies run in independent labs but negative in the ToxCast ER bioactivity model (Table 3). Due to the volatility of the chemical (157 Pa /1.18 mmHg at 25°C), it is possible that the concentration of the compound actually tested in the high-throughput assays was significantly lower that then calculated nominal concentration. The resulting overall accuracy for model performance for unequivocal chemicals was 95% (37/39; Table 4).

To broaden the evaluation of the ToxCast ER model, we compared model agonist bioactivity with 103 chemicals run in at least one guideline-like uterotrophic studies. This larger set of chemicals included the 43 *in vivo* reference chemicals as defined above, and 60 additional chemicals that did not meet the stringent criteria for an *in vivo* reference chemical because they were only run in one study or had unresolvable discordant results (*e.g.*, only one guideline-like positive and at least one guideline-like negative study for the same chemical; Figure 1). Excluding uterotrophic chemicals with equivocal ToxCast model scores left 86 chemicals, including 47 with positive uterotrophic results and 39 with only negative results (Figure 1). Of the chemicals with at least one positive uterotrophic response and unequivocal scores, 42 of 47 had ToxCast model scores $\geq$0.1. Five chemicals with positive uterotrophic studies (triclosan, reserpine, permethrin, octamethylcyclotetrasiloxane, and gibberellic acid) had no reported bioactivity in the ToxCast model (*i.e.* scores = 0). Thirty-nine chemicals did not significantly increase uterine weight in any study examined and had unequivocal ToxCast model scores. Of these, 35 had ToxCast models scores = 0 and four chemicals (phenolphthalein, benzoic acid, kaempferol, and benzylbutylphthalate) had ToxCast model scores $\geq$0.1. Though an *in vivo* response effect was not reported in the curated literature review[14], it is worth noting that phenolphthalein, kaempferol, and benzylbutylphthalate were identified as *in vitro* positive reference chemicals[31]. Concordance between the ToxCast ER model bioactivity and the *in vivo* guideline-like uterotrophic studies for the 79 chemicals with unequivocal ToxCast models scores (AUC=0 or AUC$\geq$0.1 was 89% (70/79; Table 4).

In the comparison between ToxCast model scores and Tier 1 results, three List 1 chemicals did not have ToxCast assay data and none of the remaining 49 chemicals had ToxCast ER model scores greater than 0.1. Similarly, none of the chemicals had clear positive agonist activity in the Tier 1 ER *in vitro* assays (ER binding and ERTA) or *in vivo* (uterotrophic) assays. ToxCast ER model scores were equivocal for eight List 1 Tier 1 chemicals (chemical codes 23, 22, 50, 17; 40, 28, 8, and 9, Table 5). All assay responses for these chemicals were detected at concentrations similar to those that resulted in cytotoxicity and may be explained by cell-stress or cytotoxicity-related false positive activity. Although there were both positive and negative Tier 1 ERTA assays reported for chemical codes 41 and 43, there were not clear indications of a positive Tier 1 ER binding, ERTA, or uterotrophic study (or any study submitted to EPA to satisfy a Tier 1 test order) for any chemical. Similarly ToxCast model scores were negative for the remaining 41 chemicals. Comparison between computational methods and Tier 1 assays is obviously biased by lack of positive results, but for this analysis, ToxCast model performance is 100% accurate against List 1 chemicals with Tier 1 data (Table 4).

ToxCast ER model scores were used to evaluate potential agonist activity in the ~1800 chemicals with data for all 18 *in vitro* ToxCast ER assays, including 57 of 107 List 2 chemicals. All of the 57 List 2 chemicals lacked ToxCast ER agonist bioactivity (*i.e.*, model scores = 0.1; Figure 2)[27]. However among the remaining chemicals run in ToxCast high-throughput ER assays, about 7% (133) chemicals had ToxCast ER model scores indicating positive

agonist bioactivity (*i.e.*, scores $\geq$ 0.1), 15% (276) were equivocal, and 77% (1403) have no observed ER agonist bioactivity (Figure 2)[27].

Discussion

Before new computational toxicology tools can be used for real-world applications, their utility should be adequately demonstrated for the proposed purpose. A key aspect of the analysis presented in this paper is the performance-based approach to validation using multiple sets of well-studied reference chemicals to establish the specificity and sensitivity of the ER computational model, and comparisons of model scores with existing test methods. Our analyses focused on ER agonism because the Tier 1 battery assays that measure ER interactions are only capable of detecting agonism, as it is the expected bioactivity of most estrogen-active environmental chemicals. The ToxCast ER model accurately predicted the *in vitro* bioactivity of reference compounds across a range of structures and potencies[13] and the *in vivo* ER agonist activity for a relatively large set of about 150 chemicals with bioactivities independently confirmed by another test method (*i.e.*, *in vitro* and *in vivo* reference chemicals, results of guideline-like uterotrophic studies, results of List 1/Tier 1 uterotrophic studies). Together, these analyses provide a high degree of scientific confidence in the ability of the ToxCast ER model to predict ER agonist bioactivity and demonstrate the utility of using these computational tools to meet our intended objectives: 1) to contribute to the weight of evidence evaluation of a chemical's potential ER bioactivity; and 2) to provide an alternative source of data for specific Tier 1 endpoints measuring *in vitro* and *in vivo* ER interaction.

The performance of the ER model demonstrated in this manuscript illustrates the utility of adopting a fit-for-purpose approach for validating new computational tools for use in the existing EDSP screening and testing framework. The time and resource intensive multi-laboratory approach traditionally used to validate assays, with its seven to 10 year timeline even for simple assays, is not suited to the rapid introduction of increasingly accurate and high-throughput new tools. A number of groups have proposed more rapid approaches to validation of new assays[40-45], and this case study demonstrates a key feature outlined by those authors, namely the use of performance-based validation. A performance-based validation approach including a large number of chemicals and spanning a range of structures and potencies provides scientific confidence in the validation, greatly increases our knowledge of the domain of applicability, and illustrates the strength of high-throughput assays that are capable of screening a range of chemical classes. In this study, *in vitro* reference chemicals were selected and used to test the ER model performance. The *in vitro* reference chemical activities were independently confirmed in several different types of ER assays, with reported potencies ranging five orders of magnitude. In addition, *in vivo* reference chemicals were selected and used to evaluate ER model performance for agonist bioactivity. The range of structures and potencies of the *in vivo* reference chemicals were similarly diverse and the effects (or lack of effects) on uterine weight were independently confirmed. In contrast to the large number of chemicals used in this study, the Tier 1 ER *in vitro* transactivation and binding assays were initially validated with 12 and 23 chemicals, respectively[46, 47]. Even with the relatively low number of chemicals, 35% of the results in the Tier 1 *in vitro* binding assay were not consistent with the expected outcomes, either because of lack of agreement among assay results from different labs or disagreement with observed results and anticipated activity of the selected chemicals[46]. Similarly, the OCED validation exercise of the *in vivo* uterotrophic assay relied on only seven chemicals[48].

Evaluation of performance indicates that the ToxCast ER model has 90% or greater accuracy across the *in vitro* reference chemicals, the *in vivo* reference chemicals, the larger set of chemicals with guideline-like *in vivo* uterotrophic studies, and results of Tier 1 battery assays. For use in the EDSP regulatory context, the high

sensitivity of the model is important since it means that few false negatives were observed. The performance of the ToxCast ER model is also better than lower throughput *in vitro* assays[49] which showed 66% agreement between results of competitive ER binding and uterotrophic assays weights for 65 chemicals. The superior performance of the ToxCast model was likely improved by the redundant nature of the assays targeting complementary ER signaling pathway endpoints in the ToxCast assay suite and is an improvement over the current EDSP Tier 1 assays measuring the same bioactivity.

Although comparing the ToxCast ER agonist scores with *in vivo* uterotrophic results is a critical part of the validation approach, the expectation that alternative methods will accurately predict *in vivo* responses must be tempered by the inherent variability of the *in vivo* method. Even among "guideline-like" studies, uterine response for a single chemical may vary with animal model, strain, delivery route, and dose of test chemical. Previous analysis has suggested that the relatively short duration and limited number of animals employed in the standard uterotrophic study design may result in a skewed distribution towards false negatives due to variability in uterus weights of control animals[50, 51]. Evaluation of uterotrophic study results for any single chemical often differed with animal model and delivery route used in the study, and highlights the inherent variability in uterotrophic guideline method[14, 52]. Further analysis of the full set of guideline-like uterotrophic studies indicated a moderate degree of inter-study variability. Of chemicals with >1 guideline-like studies, 26% had contradictory results with at least one positive and one negative study[14]. While much of this can be explained by dose route or animal model selected, among the 24 guideline-like uterotrophic assays conducted for bisphenol A (BPA) delivered by subcutaneous injection to the immature rat, discordant results ranged over three orders of magnitude (*e.g.,* 4 mg/kg/d produced a positive response in one study and 1000 mg/kg/d failed to do so in another; Figure 3). Comparisons between chemicals tested in both systems have led others to conclude that *in vitro* assays may be more sensitive and more reproducible than the uterine bioassay[49]. In other words, alternative methods should not be expected to predict both the true signal as well any errors associated with the *in vivo* studies.

When the performance of the ToxCast ER model is considered in the context of our stated objectives, it is clear that the ToxCast ER model has demonstrated utility for contributing to the weight of evidence of a chemical's potential interaction with the ER pathway. For the second objective of using ToxCast ER model data as an alternative for specific Tier 1 endpoints, the performance of the ToxCast ER model must also be evaluated together with the inherent variability of the existing Tier 1 assays for ER bioactivity. The 18 *in vitro* ER assays used in ToxCast include high-throughput counterparts of the Tier 1 lower throughput *in vitro*, ER binding and transactivation assays, but also include additional assays that cover other important elements of the ER signaling pathway. In this study, we describe a performance-based validation effort using a much more extensive set of *in vitro* reference chemicals than was used to validate Tier 1 *in vitro* assays and the performance of the ToxCast ER model was better than that of the existing Tier 1 ER binding and ERTA assays in their respective validation studies. Given the redundancy of coverage among the 18 ToxCast ER assays, it is unlikely that running a single guideline ER binding and/or ERTA assay would provide additional insight into a chemical's potential ER bioactivity. In addition, the more comprehensive pathway coverage and orthogonal nature of the 18 assays included in the network model provide a superior estimate of a chemical's potential ER bioactivity than existing Tier 1 ER binding and ERTA assays because the ToxCast model is capable of detecting false positives due to assay-specific interference and/or cytotoxicity that can be discriminated from "true" bioactivity[13].

Comparison of the ToxCast ER model performance with the Tier 1 uterotrophic assay also showed strong agreement. The EDSP guideline short-term uterotrophic study includes a single endpoint (uterine weight) and

as discussed above, may actually be less sensitive than *in vitro* assays[49], particularly given the redundancy in the ToxCast ER assay suite that is likely to reduce false negative responses. While the current high-throughput ER assays have limited capacity to address chemicals that may be biotransformed to active or inactive metabolites, the preferred method of administration for the guideline uterotrophic assay is subcutaneous injection which bypasses first pass hepatic metabolism. The performance-based evaluation presented in this paper makes a strong case for using the ToxCast ER model as an alternative for the Tier 1 uterotrophic assay. It should be noted that these analyses do not support substitution of all EDSP Tier 1 assays that provide information on potential estrogen pathway interactions. In the absence of an *in vivo* uterotrophic assay, other estrogen-pathway Tier 1 *in vivo* assays, such as the rat pubertal assay[53], which includes a variety of endpoints in addition to uterine weight, could detect chemicals that are metabolized to active metabolites through the oral administration of test chemical. In addition, the rat pubertal assay exposes a larger sample of animals (n=16 versus n=6) for a longer duration (20 days versus three days), which may increase the sensitivity of the assay compared with the uterotrophic test. In light of this examination, if ToxCast *in vitro* assay data were available for a given chemical, no additional information on potential ER agonist bioactivity would be gained by requiring a Tier 1 uterotrophic assay.

Lastly, the EDSP Tier 1 battery is intended to screen potential endocrine effects in humans *and* wildlife. ToxCast *in vitro* ER assays measure effects using cells derived from the kidney, cervix, liver, ovary, uterus, and breast; use rodent, bovine, and human receptor proteins; and detect interaction using a variety of technologies (Table 1). This diversity of the 18 high-throughput ER assays accounts for estrogenic effects more broadly across cell types, organs, and species than the single human ovarian and two rodent uterine ER assays in the existing Tier 1 battery and may have greater relevance to wildlife.

We have demonstrated the performance of the ToxCast ER model for predicting ER bioactivity of *in vitro* and *in vivo* reference chemicals, the utility of using the ToxCast ER model bioactivity as an alternative to the EDSP Tier 1 ER binding, ERTA, and uterotrophic endpoints, and using ToxCast ER model scores to prioritize chemicals in the EDSP universe for additional screening and testing. Results of the ER model indicate only about 7% of the 1800 chemicals run in ToxCast have potential significant ER agonist bioactivity and this subset does not include any List 1 or List 2 chemicals. The lack of ToxCast model ER bioactivity among EDSP List 1 chemicals is consistent with the absence of ER activity observed in the initial review of Tier 1 battery data for a subset of 21 of List 1 chemicals[54, 55]. The ToxCast ER model can be used to rapidly screen the ~10,000 chemicals in the EDSP universe, allow EPA to move away from screening lists of few chemicals with relatively low or no potential endocrine activity, and identify chemicals with the greatest potential endocrine bioactivity that may be high priority candidates for further screening and testing[54, 55]. This approach for using computational toxicology tools in the EDSP only evaluated ER-mediated bioactivity and it should be noted that while List 1 and List 2 chemicals appear to have limited ER bioactivity, these chemicals may be active in other endocrine pathways. In the future, we plan to use a performance-based validation approach of high-throughput ToxCast ER assays and other computational toxicology tools to compare with the existing "guideline-like" Tier 1 assays, including the fish and pubertal rat, to determine how well high-throughput models predict estrogen bioactivity in neuroendocrine-intact animals. In addition, EPA will use this performance-based approach for validating new computational tools to screen for androgen and thyroid effects, taking advantage of both existing and innovative, emerging technologies to implement a scientifically robust and comprehensive chemical prioritization process for EDSP.

Acknowledgements

We thank D. Bergfelt, A. Dixon, T. Green, K. Hamernik, S. Lynn, K. Markey, L. Touart, and W. Wooge from OSCP; K. Houck, K. Crofton, T. Knudsen, M. Martin, A. Richard, W. Setzer, I. Shah, J. Wambaugh, T. Bahadori from ORD for their contributions. We are grateful to the NICEATM/ILS team for curation of the uterotrophic database. We also thank R. Hines from ORD for his technical review of the manuscript.

Table 1. Summary of the 18 high-throughput *in vitro* estrogen receptor (ER) assays included in the ToxCast ER bioactivity model. For additional information on assays, please reference the EDSP21 dashboard[56]. NA = Not Applicable (cell-free binding assay).

Assay ID	Assay Name	Biological Process Target	Detection Technology	Organism	Tissue	Cell Line
A1	NVS_NR_bER	receptor binding	radioligand	bovine	uterus	NA
A2	NVS_NR_hER	receptor binding	radioligand	human	NA	NA
A3	NVS_NR_mERa	receptor binding	radioligand	mouse	NA	NA
A4	OT_ER_ERaERa_0480	protein complementation	fluorescence	human	kidney	HEK293T
A5	OT_ER_ERaERa_1440	protein complementation	fluorescence	human	kidney	HEK293T
A6	OT_ER_ERaERb_0480	protein complementation	fluorescence	human	kidney	HEK293T
A7	OT_ER_ERaERb_1440	protein complementation	fluorescence	human	kidney	HEK293T
A8	OT_ER_ERbERb_0480	protein complementation	fluorescence	human	kidney	HEK293T
A9	OT_ER_ERbERb_1440	protein complementation	fluorescence	human	kidney	HEK293T
A10	OT_ERa_EREGFP_0120	gene expression	fluorescence	human	cervix	HeLa
A11	OT_ERa_EREGFP_0480	gene expression	fluorescence	human	cervix	HeLa
A12	ATG_ERa_TRANS_up	mRNA induction	fluorescence	human	liver	HepG2
A13	ATG_ERE_CIS_up	mRNA induction	fluorescence	human	liver	HepG2
A14	Tox21_ERa_BLA_Agonist_ratio	gene expression	fluorescence	human	kidney	HEK293T
A15	Tox21_ERa_LUC_BG1_Agonist	gene expression	bioluminescence	human	ovary	BG1
A16	ACEA_T47D_80hr_Positive	cell proliferation	electrical impedance	human	breast	T47D
A17	Tox21_ERa_BLA_Antagonist_ratio	gene expression	fluorescence	human	kidney	HEK293T
A18	Tox21_ERa_LUC_BG1_Antagonist	gene expression	bioluminescence	human	ovary	BG1

Table 2: *In vitro* estrogen receptor (ER) agonist reference chemicals.

CASRN	Chemical Name	Agonist Potency[1]	ToxCast ER Model Score
57-63-6	17alpha-Ethinyl estradiol	Strong	1
84-16-2	meso-Hexestrol	Strong	0.99
56-53-1	Diethylstilbestrol (DES)	Strong	0.94
50-28-2	17beta-Estradiol	Strong	0.94
57-91-0	17alpha-Estradiol	Moderate	1.06
53-16-7	Estrone	Moderate	0.81
140-66-9	4-tert-Octylphenol	Moderate	0.39
446-72-0	Genistein	Weak	0.54
77-40-7	Bisphenol B	Weak	0.49
80-05-7	Bisphenol A	Weak	0.45
486-66-8	Daidzein	Weak	0.44
521-18-6	5alpha-Dihydrotestosterone	Weak	0.40
789-02-6	o,p'-DDT	Weak	0.39
599-64-4	4-Cumylphenol	Weak	0.38
143-50-0	Kepone	Weak	0.17
58-18-4	17alpha-Methyltestosterone	Very Weak	0.50
520-36-5	Apigenin	Very Weak	0.31
72-43-5	Methoxychlor	Very Weak	0.25
520-18-3	Kaempferol	Very Weak	0.25
85-68-7	Butylbenzyl phthalate	Very Weak	0.18
480-40-0	Chrysin	Very Weak	0.13
60168-88-9	Fenarimol	Very Weak	0.11
104-40-5	p-n-Nonylphenol	Very Weak	0.1
120-47-8	Ethylparaben	Very Weak	0.1
72-55-9	p,p'-DDE	Very Weak	0.1
84-74-2	Di-n-butyl phthalate	Very Weak	0.03
115-32-2	Dicofol	Very Weak	0
117-81-7	Diethylhexyl phthalate	Very Weak	0
52-86-8	Haloperidol	Inactive	0.01
52-01-7	Spironolactone	Inactive	0
50-22-6	Corticosterone	Inactive	0
13311-84-7	Flutamide	Inactive	0
1912-24-9	Atrazine	Inactive	0
32809-16-8	Procymidone	Inactive	0
330-55-2	Linuron	Inactive	0
50-55-5	Reserpine	Inactive	0
52806-53-8	Hydroxyflutamide	Inactive	0
57-30-7	Phenobarbital Sodium	Inactive	0
65277-42-1	Ketoconazole	Inactive	0
66-81-9	Cycloheximide	Inactive	0

[1]Strong = AC_{50} <0.0001 µM, moderate = AC_{50} < 0.1 µM, weak = AC_{50} < 1 µM, very weak = all other activities, inactive = no detected activity.

Table 3: *In vivo* estrogen receptor (ER) agonist reference chemicals with at least two independent active or inactive guideline-like uterotrophic studies identified from a comprehensive literature review and curation (reproduced from Kleinstreuer *et al.*[14] with permission). The numbers of guideline-like active and inactive study results are reported for each chemical.

CASRN	Name	Active	Inactive	Bioactivity	ToxCast ER Model Score
57-91-0	17alpha-Estradiol	2	0	Active	1.06
57-63-6	Ethinyl Estradiol	59	0	Active	1
56-53-1	Diethylstilbestrol (DES)	8	1	Active	0.94
50-28-2	Estradiol	25	0	Active	0.94
474-86-2	Equilin	2	0	Active	0.82
53-16-7	Estrone	9	0	Active	0.81
50-27-1	Estriol	4	0	Active	0.79
72-33-3	Mestranol	3	0	Active	0.74
17924-92-4	Zearalenone	4	0	Active	0.71
1478-61-1	Bisphenol AF	4	0	Active	0.55
446-72-0	Genistein	27	1	Active	0.54
68-22-4	Norethindrone	2	0	Active	0.52
58-18-4	Methyltestosterone	3	0	Active	0.50
77-40-7	Bisphenol B	2	0	Active	0.49
80-05-7	Bisphenol A	37	6	Active	0.45
104-43-8	4-Dodecylphenol	3	0	Active	0.41
521-18-6	Dihydrotestosterone	3	0	Active	0.4
131-55-5	Benzophenone-2	6	0	Active	0.40
140-66-9	4-(1,1,3,3-Tetramethylbutyl)phenol	3	1	Active	0.39
789-02-6	o,p'-DDT	15	1	Active	0.39
599-64-4	p-Cumylphenol	2	0	Active	0.38
5153-25-3	Benzoic acid, 4-hydroxy-, 2-ethylhexyl ester	2	0	Active	0.37
80-46-6	4-(1,1-Dimethylpropyl)phenol	4	0	Active	0.28
131-56-6	2,4-Dihydroxybenzophenone	3	0	Active	0.27
80-09-1	Bisphenol S	2	0	Active	0.26
72-43-5	Methoxychlor	18	1	Active	0.25
94-26-8	Butylparaben	8	2	Active	0.25
98-54-4	p-tert-Butylphenol	2	0	Active	0.16
104-40-5	Nonylphenol	5	4	Active	0.10
556-67-2	Octamethylcyclotetrasiloxane	3	0	Active	0
520-18-3	Kaempferol	0	3	Inactive	0.25
84-74-2	Dibutyl phthalate	0	2	Inactive	0.03
84-61-7	Dicyclohexyl phthalate	0	2	Inactive	0.02
84-75-3	Dihexyl phthalate	0	2	Inactive	0.01
51630-58-1	Fenvalerate	0	2	Inactive	0.01
103-23-1	Bis(2-ethylhexyl)hexanedioate	0	2	Inactive	0

117-81-7	Bis(2-ethylhexyl)phthalate	0	2	Inactive	0
1461-22-9	Tributylchlorostannane	0	2	Inactive	0
1912-24-9	Atrazine	0	2	Inactive	0
61-82-5	Amitrole	0	2	Inactive	0
84-66-2	Diethyl phthalate	0	2	Inactive	0
87-86-5	Pentachlorophenol	0	2	Inactive	0
99-96-7	4-Hydroxybenzoic acid	0	2	Inactive	0

Table 4. Performance based evaluation of the ToxCast ER model based on 18 high-throughput *in vitro* assays measuring potential estrogen receptor (ER) agonist activities and *in vitro* reference chemicals (see text for detailed explanation). ToxCast ER model scores ≥0.1 were considered positive, negative scores = 0 (and values <0.001 were truncated as 0). For the purposes of these analyses equivocal model scores (0>AUC<0.1) were excluded.

Performance	*In vitro* reference chemicals	*In vivo* reference chemicals	GL uterotrophic studies	Tier 1 studies
# True Pos	25	29	42	0
# True Neg	11	8	35	41
# False Pos	0	1	4	0
# False Neg	2	1	5	41
Accuracy	0.95	0.95	0.90	1.0
Sensitivity	0.93	0.97	0.89	0
Specificity	1.0	0.89	0.90	1.0

Table 5. Estrogen receptor (ER) agonist activity of List 1 chemicals determined from ToxCast ER model scores for agonist bioactivity compared to results of EDSP Tier 1 (T1) estrogen receptor binding, estrogen receptor transactivation (ERTA), and uterotrophic (UT) assay results. OSRI= other scientifically relevant information; N = negative; P = positive; Incl = inconclusive; Equiv = equivocal.

Code[1]	ToxCast ER Model[2]	T1 ER binding[3]	T1 ERTA[3]	T1 UT[3]
23	0.04	N	19.10%	N
22	0.03	N	15.80%	N
50	0.03	N	OSRI (P)	N
17	0.02	Equiv	N	N
40	0.01	OSRI (N)	N	N
28	0.01	N	18.40%	N
8	0.01	Equiv	N	N
9	0.01	Equiv	OSRI (P)	N
38	0	N	24.70%	OSRI (N)
5	0	OSRI (N)	OSRI (N)	OSRI (N)
11	0	N	Incl	N
39	0	N	N	N
48	0	N	Incl	N
43	0	N	OSRI (P/N)	N
4	0	OSRI (N)	N	OSRI (N)
6	0	N	N	OSRI (N)
7	0	N	N	N
10	0	OSRI (N)	N	N
12	0	N	Incl	N
13	0	OSRI (N)	Incl	OSRI (N)
14	0	N	Incl	N
15	0	N	Incl	N
16	0	N	OSRI (N)	N
17	0	N	N	N
19	0	Incl	N	N
20	0	N	N	N
21	0	N	Incl	N
24	0	Equiv	Incl	N
25	0	N	Incl	N
26	0	OSRI (N)	OSRI (N)	OSRI (N)
27	0	N	N	N
29	0	N	N	N
30	0	N	Incl	N
31	0	N	N	N

32	0	N	Incl	N
33	0	N	Incl	N
34	0	N	N	N
35	0	N	Incl	N
36	0	N	N	N
37	0	N	N	N
41	0	OSRI (N)	OSRI (P/N)	OSRI (N)
42	0	Equiv	N	N
44	0	OSRI (N)	N	OSRI (N)
45	0	N	OSRI (N)	N
46	0	N	N	N
47	0	N	N	N
49	0	N	N	N
51	0	OSRI (N)	N	N
52	0	N	N	N
1	NA	N	N	N
2	NA	N	N	N
3	NA	N	Incl	N

[1]Note: List 1 chemicals in the table below are coded because Tier 1 weight of evidence determinations for List 1 chemicals are not final.

[2]ToxCast results for three List 1 chemicals were not available (NA).

[3]In some cases, Tier 1 test requirements were satisfied by other scientifically relevant information (OSRI) such as a similar study published in peer-reviewed literature.

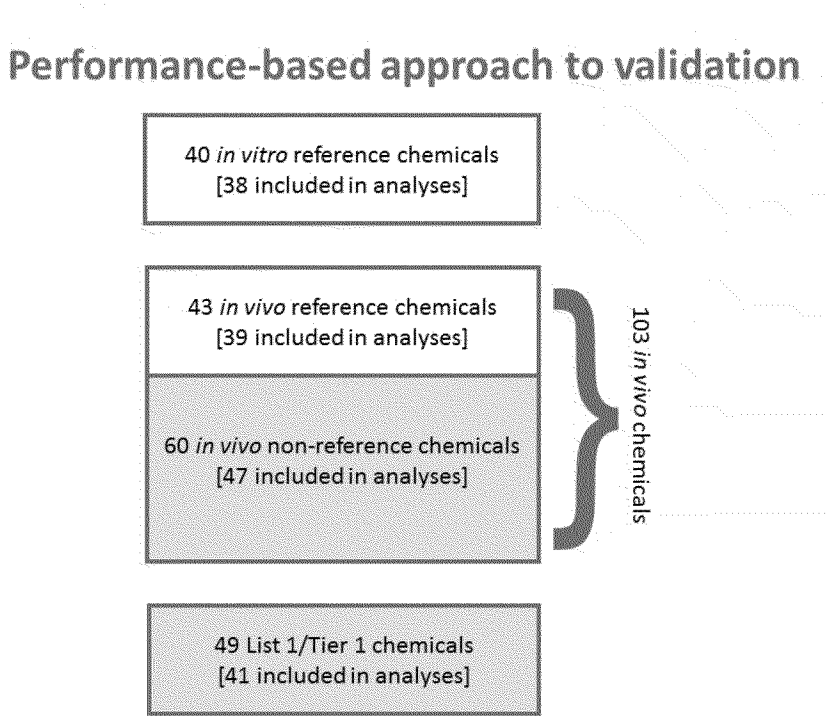

Figure 1. Diagram of the different chemical sets used in the performance-based approach to establishing scientific confidence in the ToxCast ER model for determining bioactivity of test chemicals. *In vitro* reference chemicals were selected from ICCVAM and OECD published validations of *in vitro* assays. "Guideline-like UT" refers to 103 chemicals tested in uterotrophic studies published in peer-reviewed literature and determined to be "guideline-like" based on methodological consistency with EDSP Tier 1 uterotrophic guidelines. A subset of these chemicals had uterotrophic results that were independently confirmed in at least two guideline studies. For 49 (of 52) EDSP List 1 chemicals, ToxCast ER model results were compared with results of in EPA guideline estrogen receptor (ER) binding, ER transactivation (ERTA), and uterotrophic studies. UT= uterotrophic assay. ToxCast ER model scores indicating bioactivity (scores>0.1; approximating an AC_{50} of 100 µM) or no activity (scores=0) were included in analyses. Equivocal model scores (<0.1 and >0) were excluded because putative *in vitro* activity in this range was generally beyond the limit of detection. See text for details.

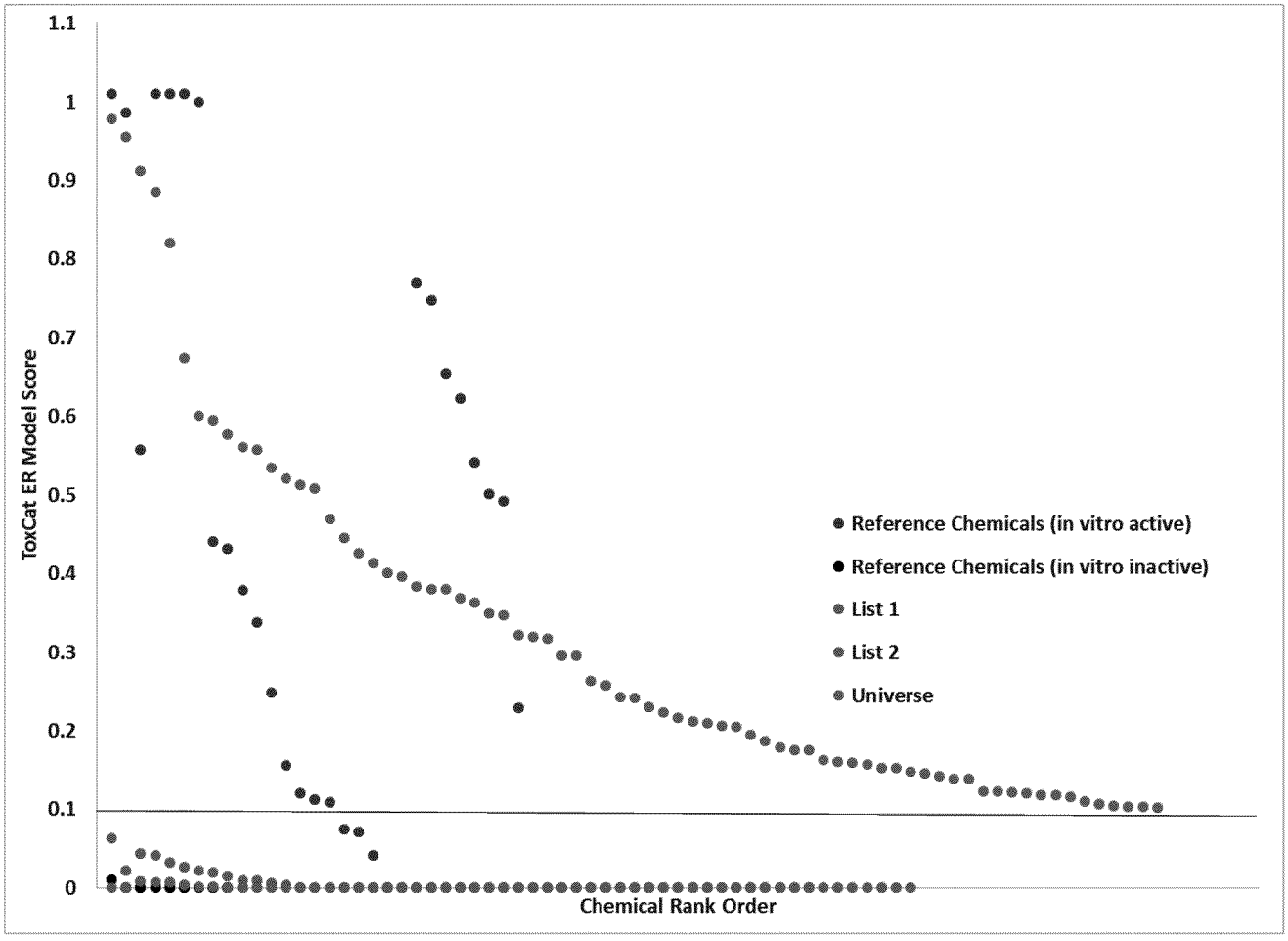

Figure 2. ToxCast ER model agonist bioactivity scores for EDSP List 1, List 2, Universe, and reference chemicals. Scores greater than 0.1 (indicated by the horizontal line) were considered positive. Scores for 40 active (green) and inactive (black) *in vitro* reference chemicals are shown, along with scores for List 1 (blue) and List 2 (orange) chemicals run in ToxCast. Of the approximately 1800 chemical run chemicals run in ToxCast high-throughput ER assays, about 7% (133) have ER ToxCast model scores that indicate potential agonist bioactivity (scores $\geq$ 0.1) and 77% (1403) are negative (scores = 0; values truncated <0.001).

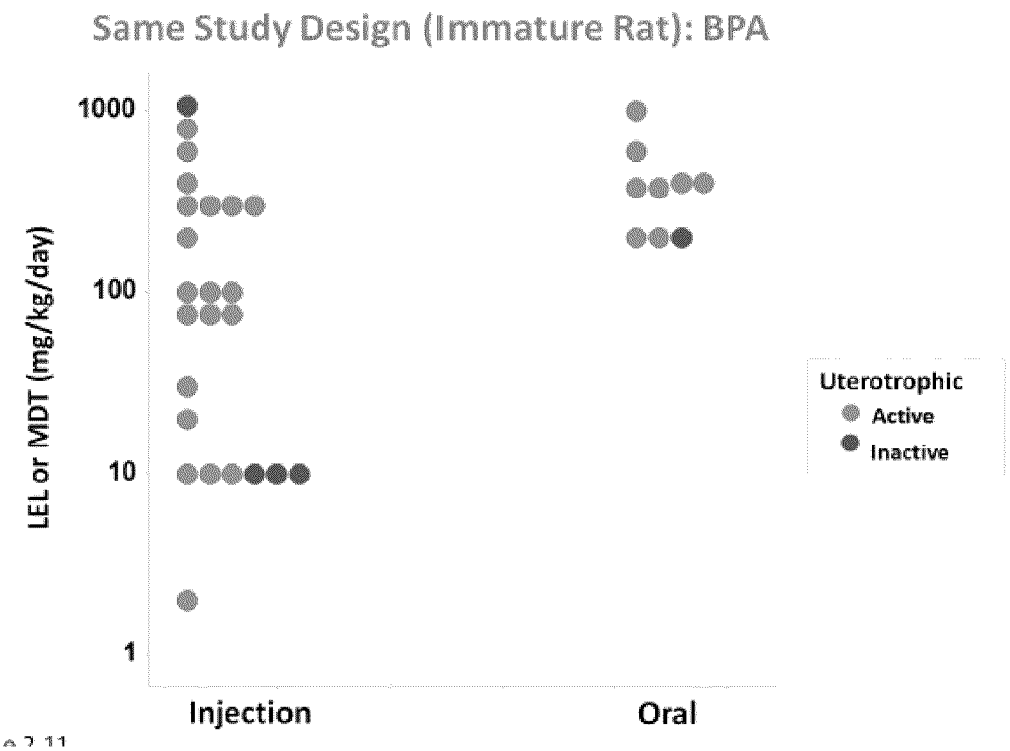

Figure 3. An illustration of the variability of results of bisphenol A (BPA) uterotrophic studies conducted in the immature rat model. All studies summarized below are methodologically similar to the EDSP Tier 1 study and are considered "guideline-like", yet have discordant results even within for the same route of administration. LEL = lowest effect level; MDT = maximum dose tested.

Literature Cited

1. U.S. EPA EDSP Tier 1 Battery of Assays http://www.epa.gov/endo/pubs/assayvalidation/tier1battery.htm
2. U.S. EPA EDSP Overview. http://www.epa.gov/endo/pubs/edspoverview/components.htm
3. U.S. EPA Final List of Chemicals for Initial Tier 1 Screening. http://www.epa.gov/endo/pubs/prioritysetting/finallist.html
4. U.S. EPA Revised Second List of Chemicals for Tier 1 Screening. http://www.epa.gov/endo/pubs/prioritysetting/revlist2.htm
5. U.S. EPA Universe of Chemicals and General Validation Principles. http://www.epa.gov/endo/pubs/edsp_chemical_universe_and_general_validations_white_paper_11_12.pdf
6. NRC, *Toxicity testing in the 21st century: A vision and a strategy*. National Academies Press: Washington, DC, 2007; Vol. 25, p 136-138.
7. U.S. EPA FY 2012 EPA Budget In Brief. http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100A5RE.txt
8. U.S. EPA The U.S. Environmental Protection Agency's Strategic Plan for Evaluating the Toxicity of Chemicals http://www.epa.gov/spc/toxicitytesting/docs/toxtest_strategy_032309.pdf
9. U.S. EPA Endocrine Disruptor Screening Program Universe of Chemicals and General Validation Principles. http://www.epa.gov/endo/pubs/edsp_chemical_universe_and_general_validations_white_paper_11_12.pdf
10. Kojima, H.; Katsura, E.; Takeuchi, S.; Niyama, K.; Kobayashi, K., Screening for estrogen and androgen receptor activities in 200 pesticides by *in vitro* reporter gene assays using Chinese hamster ovary cells. *Environmental health perspectives* **2004,** *112*, (5), 524-31.
11. Safe, S. H.; Pallaroni, L.; Yoon, K.; Gaido, K.; Ross, S.; Saville, B.; McDonnellc, D., Toxicology of environmental estrogens. *Reproduction, fertility, and development* **2001,** *13*, (4), 307-15.
12. U.S. EPA EPA's ToxCast Program. http://www.epa.gov/ncct/toxcast/
13. Judson, R. S.; Magpantay, F. M.; Chickarmane, V.; Haskell, C.; Tania, N.; Taylor, J.; Xia, M.; Huang, R.; Rotroff, D. M.; Filer, D. M.; Houck, K. A.; Martin, M. T.; Sipes, N.; Richard, A. M.; Mansour, K.; Setzer, R. W.; Knudsen, T. B.; Crofton, K. M.; Thomas, R. S., Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High Throughput Screening Assays for the Estrogen Receptor. In 2015.
14. Kleinstreuer, N. C.; Ceger, P.; Allen, D.; Strickland, J.; Chang, X.; Hamm, J.; Casey, W., A Curated Database of Rodent Uterotrophic Bioactivity. In 2015.
15. Dix, D. J.; Houck, K. A.; Martin, M. T.; Richard, A. M.; Setzer, R. W.; Kavlock, R. J., The ToxCast Program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* **2007,** *95*, 7.
16. Houck, K. A.; Kavlock, R. J., Understanding mechanisms of toxicity: Insights from drug discovery research. *Toxicology and Applied Pharmacology* **2008,** *227*, 15.
17. Martin, M. T.; Knudsen, T. B.; Reif, D. M.; Houck, K. A.; Judson, R. S.; Kavlock, R. J.; Dix, D. J., Predictive Model of Rat Reprodcutive Toxicity from ToxCast High Throughput Screening *Biology of Reproduction* **2011,** *85*, (2), 12.
18. Rotroff, D. M.; Dix, D. J.; Houck, K. A.; Knudsen, T. B.; Martin, M. T.; McLaurin, K. W.; Reif, D. M.; Crofton, K. M.; Singh, A. V.; Xia, M. H.; Huang, R. L.; Judson, R. S., Using in Vitro High Throughput Screening Assays to Identifiy Potential Endocrine Disrupting Chemicals. *Environmental health perspectives* **2013,** *121*, 7.
19. Reif, D. M.; Martin, M. T.; Tan, S. W.; Houck, K. A.; Judson, R. S.; Richard, A. M.; Knudsen, T. B.; Dix, D. J.; Kavlock, R. J., Endocrine Profile and Prioritization of Environmental Chemicals Using ToxCast Data. *Environmental health perspectives* **2010,** *118*, (12), 6.
20. Dix, D. J.; Houck, K. A.; Martin, M. T.; Richard, A. M.; Setzer, R. W.; Kavlock, R. J., The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* **2007,** *95*, (1), 5-12.
21. Judson, R. S.; Houck, K. A.; Kavlock, R. J.; Knudsen, T. B.; Martin, M. T.; Mortensen, H. M.; Reif, D. M.; Rotroff, D. M.; Shah, I.; Richard, A. M.; Dix, D. J., In vitro screening of environmental chemicals for targeted

testing prioritization: the ToxCast project. *Environmental health perspectives* **2010,** *118*, (4), 485-92.
22. Knudsen, T.; Houck, K.; Judson, R. S.; Singh, A.; Mortensen, H.; Reif, D.; Dix, D. J.; Kavlock, R. J., Biochemical Activities of 320 ToxCast Chemicals Evaluated Across 239 Functional Targets. *Toxicology* **2011,** *282*, (1-2), 1-15.
23. Sipes, N. S.; Martin, M. T.; Kothiya, P.; Reif, D. M.; Judson, R. S.; Richard, A. M.; Houck, K. A.; Dix, D. J.; Kavlock, R. J.; Knudsen, T. B., Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. *Chemical research in toxicology* **2013,** *26*, (6), 878-95.
24. Martin, M. T.; Dix, D. J.; Judson, R. S.; Kavlock, R. J.; Reif, D. M.; Richard, A. M.; Rotroff, D. M.; Romanov, S.; Medvedev, A.; Poltoratskaya, N.; Gambarian, M.; Moeser, M.; Makarov, S. S.; Houck, K. A., Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. *Chemical research in toxicology* **2010,** *23*, (3), 578-90.
25. Huang, R.; Sakamuru, S.; Martin, M. T.; Reif, D. M.; Judson, R. S.; Houck, K. A.; Casey, W.; Hsieh, J.-H.; Shockley, K. R.; Ceger, P.; Fostel, J.; Witt, K. L.; Tong, W.; Rotroff, D. M.; Zhao, T.; Shinn, P.; Simeonov, A.; Dix, D. J.; Austin, C. P.; Kavlock, R. J.; Tice, R. R.; Xia, M., Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. *Scientific Reports* **2014,** *4*.
26. Rotroff, D. M.; Dix, D. J.; Houck, K. A.; Kavlock, R. J.; Knudsen, T. B.; Martin, M. T.; Reif, D. M.; Richard, A. M.; Sipes, N. S.; Abassi, Y. A.; Jin, C.; Stampfl, M.; Judson, R. S., Real-Time Growth Kinetics Measuring Hormone Mimicry for ToxCast Chemicals in T-47D Human Ductal Carcinoma Cells. *Chemical research in toxicology* **2013**.
27. U.S. EPA U.S. EPA ToxCast Program Data. http://epa.gov/ncct/toxcast/data.htm
28. Filer, D. The ToxCast Analysis Pipeline: An R Package for Processing And Modeling Chemical Screening Data. http://www.epa.gov/ncct/toxcast/files/MySQL%20Database/Pipeline_Overview.pdf
29. U.S. EPA Integrated Bioactivity and Exposure Ranking: A computational Approach for the Prioritization and Screening of Chemicals in the Endocrine Disruptor Screening Program. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2014-0614-0003
30. OECD Test No. 455: Performance-Based Test Guideline for Stably Transfected Transactivation In Vitro Assays to Detect Estrogen Receptor Agonists. http://www.oecd-ilibrary.org/environment/test-no-455-performance-based-test-guideline-for-stably-transfected-transactivation-in-vitro-assays-to-detect-estrogen-receptor-agonists_9789264185388-en
31. ICCVAM The LUMI-CELL® ER (BG1Luc ER TA) Test Method: An In Vitro Assay for Identifying Human Estrogen Receptor Agonist and Antagonist Activity of Chemicals. http://ntp.niehs.nih.gov/pubhealth/evalatm/iccvam/test-method-evaluations/endocrine-disruptors/bg1luc/tmer/index.html
32. OECD OECD Test No. 457: BG1Luc Estrogen Receptor Transactivation Test Method for Identifying Estrogen Receptor Agonists and Antagonists. http://www.oecd-ilibrary.org/environment/test-no-457-bg1luc-estrogen-receptor-transactivation-test-method-for-identifying-estrogen-receptor-agonists-and-antagonists_9789264185395-en
33. OECD Test No. 440: Uterotrophic Bioassay in Rodents A short-term screening test for oestrogenic properties http://www.oecd-ilibrary.org/environment/test-no-440-uterotrophic-bioassay-in-rodents_9789264067417-en
34. U.S. EPA Endocrine Disruptor Screening Program Test Guidelines OPPTS 890. 1600: Uterotrophic Assay. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0012
35. U.S. EPA EPA ACToR. http://actor.epa.gov/actor
36. FDA, U. S. US Food and Drug Administration's Endocrine Disruptor Knowledge Base. http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/default.htm
37. U.S. EPA Endocrine Disruptor Screening Program Test Guidelines OPPTS 890. 1250: Rat Uterine Cytosol Estrogen Receptor Binding Assay. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0005

38. U.S. EPA, Endocrine Disruptor Screening Program Test Guidelines OPPTS 890. 1300: Estrogen Receptor Transactivation Assay. In 2009.
39. Zava, D. T., Estrogenic and antiproliferative properties of genestine and other flavonoids in human breast cancer cells *in vitro*. *Nutrition and Cancer* **1997,** *27*, (1), 31-40.
40. Becker, R. A.; Friedman, K. P.; Simon, T. W.; Marty, M. S.; Patlewicz, G.; Rowlands, J. C., An exposure:activity profiling method for interpreting high-throughput screening data for estrogenic activity--proof of concept. *Regulatory toxicology and pharmacology : RTP* **2015,** *71*, (3), 398-408.
41. Patlewicz, G.; Ball, N.; Boogaard, P. J.; Becker, R. A.; Hubesch, B., Building scientific confidence in the development and evaluation of read-across. *Regulatory toxicology and pharmacology : RTP* **2015,** *72*, (1), 117-33.
42. Patlewicz, G.; Simon, T. W.; Rowlands, J. C.; Budinsky, R. A.; Becker, R. A., Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. *Regulatory toxicology and pharmacology : RTP* **2015,** *71*, (3), 463-77.
43. Judson, R.; Kavlock, R.; Martin, M.; Reif, D.; Houck, K.; Knudsen, T.; Richard, A.; Tice, R. R.; Whelan, M.; Xia, M.; Huang, R.; Austin, C.; Daston, G.; Hartung, T.; Fowle, J. R., 3rd; Wooge, W.; Tong, W.; Dix, D., Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *Altex* **2013,** *30*, (1), 51-6.
44. Sauer, J. M.; Hartung, T.; Leist, M.; Knudsen, T. B.; Hoeng, J.; Hayes, A. W., Systems Toxicology: The Future of Risk Assessment. *International journal of toxicology* **2015**.
45. Tralau, T.; Oelgeschlager, M.; Gurtler, R.; Heinemeyer, G.; Herzler, M.; Hofer, T.; Itter, H.; Kuhl, T.; Lange, N.; Lorenz, N.; Muller-Graf, C.; Pabel, U.; Pirow, R.; Ritz, V.; Schafft, H.; Schneider, H.; Schulz, T.; Schumacher, D.; Zellmer, S.; Fleur-Bol, G.; Greiner, M.; Lahrssen-Wiederholt, M.; Lampen, A.; Luch, A.; Schonfelder, G.; Solecki, R.; Wittkowski, R.; Hensel, A., Regulatory toxicology in the twenty-first century: challenges, perspectives and possible solutions. *Arch Toxicol* **2015,** *89*, (6), 823-50.
46. U.S. EPA Integrated Summary Report for Validation of an Estrogen Receptor Binding Assay Using Rat Uterine Cytosol as Source of Receptor as a Potential Screen in the Endocrine Disruptor Screening Program Tier 1 Battery. http://www.epa.gov/endo/pubs/assayvalidation/status.htm
47. OECD Draft Report of Pre-validation and Inter-laboratory Validation For Stably Transfected Transcriptional Activation (TA) Assay to Detect Estrogenic Activity - The Human Estrogen Receptor Alpha Mediated Reporter Gene Assay Using hER-HeLa-9903 Cell Line.
http://www.oecd.org/chemicalsafety/testing/37504278.pdf
48. OECD Series on Testing and Assessment. Number 67. Report of the Validation of the Uterotrophic Bioassay: Additional Data Supporting the TEst Guideline on the Uterotrophic Bioassay in Rodents. .
http://epa.gov/endo/pubs/uterotrophic_OECD_validation_report.pdf
49. Akahori, Y.; Nakai, M.; Yamasaki, K.; Takatsuki, M.; Shimohigashi, Y.; Ohtaki, M., Relationship Between the Results of *in vitro* Receptor Binding Assay to Human Estrogen Receptor α and *in vivo* Uterotrophic Assay: Comparative Study with 65 Selected Chemicals. *Toxicology in Vitro* **2008,** *22*, 7.
50. Ashby, J.; Odum, J., Gene expression changes in the immature rat uterus: effects of uterotrophic and sub-uterotrophic doses of bisphenol A. *Toxicological Sciences* **2004,** *82*, (2), 458-467.
51. Christian, M. S.; Hoberman, A. M.; Bachmann, S.; Hellwig, J., Variability in the uterotrophic response assay (an in vivo estrogenic response assay) in untreated control and positive control (DES-DP, 2.5 microG/kg, bid) Wistar and Sprague-Dawley rats. . *Drug and Chemical Toxicology* **1998,** *21 Supple 1*, 50.
52. Laws, S. C.; Carey, S. A.; Ferrell, J. M.; Bodman, G. J.; Cooper, R. L., Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats. *Toxicological sciences : an official journal of the Society of Toxicology* **2000,** *54*, (1), 154-67.
53. U.S. EPA Endocrine Disruptor Screening Program Test Guidelines: OPPTS 890.1450. Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Female Rats.

http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0009
54. U.S. EPA SAP Minutes No: 2013-05 - Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening. http://www.epa.gov/scipoly/sap/meetings/2013/july/073013minutes.pdf
55. U.S. EPA Final EDSP (Endocrine Disruptor Screening Program) White Paper on Tier I Assay and Battery Performance for May 21-23, 2013 FIFRA SAP Meeting. http://www.regulations.gov/contentStreamer?objectId=09000064812a3e5d&disposition=attachment&contentType=pdf
56. U.S. EPA EDSP21 Dashboard. http://actor.epa.gov/edsp21/